# Dundee Discussion Papers in Economics 283

Jones, Martin

*Publication date:*
2015

# Dundee Discussion Papers in Economics


# Values, Multiculturalism and Representations


Martin K Jones



Department of
Economic Studies
University of Dundee
Dundee
DD1 4HN



Working Paper
No. 283
January 2015

ISSN: 1473-236X

Values, Multiculturalism and Representations

Martin K Jones*

Abstract: The theory of values is underdeveloped within economics with very few theoretical models explaining how values are created and propagated through a population. Those models that do exist are limited in how well they explain the role of values and the conclusions they reach. This paper argues that the theory of values needs to be fundamentally rethought, incorporating basic ideas from psychology, anthropology and philosophy that have largely been ignored in the literature. The resulting models are applied to answer some fundamental issues surrounding the notion of multiculturalism.

JEL Classification:

Keywords: Values, Representations, Mental Actions, Multiculturalism

*Economic Studies, School of Social Sciences, University of Dundee, Dundee. DD1 4HN.

M.K.Jones@dundee.ac.uk tel: +44 1382 385164 fax: +44 1382 384691

## 1. Introduction

The influence of culture and, in particular, cultural values is an active and promising area of research within economics, resulting in a variety of field and experimental studies in the literature (See de Jong (2009) for an overview). It is becoming increasingly obvious that values have a huge impact on economic and other behaviour. Different values within a society can lead to political disagreements and clashes, different business values lead to different employment practices, different consumer values lead to different goods being bought and different values between societies can have a serious effect on the ability to trade and invest in different countries.

However, the theoretical literature has lagged behind with few papers actively trying to model *values* as distinct from norms and conventions. Within cultural economics, values are generally instrumentalised from questionnaires that are in widespread use in anthropology and sociology (See de Jong (2009) and Inglehart et al. (1998) for examples) and have little theoretical foundation. This is a shame in that a reasonable theory of values would help to explain some of the culture clashes that do occur and allow us to predict which values are likely to emerge within cultures. Furthermore, the papers that do exist do not fit the notion of values very well in that they ignore certain common characteristics that should be explained by any reasonable theory. It will be argued that to explain these characteristics requires us to take seriously concepts that are widely used in cognitive disciplines but are not widely used in economics.

In order to demonstrate the applicability of these ideas we will construct a model of values that casts some light on the issue of multiculturalism and the issues brought up by multiculturalism. Culture consists of many things: objects, icons but also , to a large extent, values (see Sperber 1996). This means that a theoretical study of values may have something to tell us about how likely particular cultures are to survive in competition with each other. Is a multicultural society likely to survive or are some cultures going to be assimilated by others? How crucial are institutions within society in the preservation of multiculturalism? The simple models presented here will give some preliminary answers to these questions.

## 2.Other Theories of Value

Values are quite often ignored in economics even when they are the ostensible subject of discussion. An example of this is in the book "Economics, Values and Organization" (1998) edited by Ben-ner and Putterman. In this book are papers by a wide variety of academics from various disciplines. What is interesting is that among the economists (e.g. Robert Sugden, Ken Binmore, Timur Kuran) values, as such, are only mentioned in passing or they are conflated with other moral concepts such as norms. An example is in Sugden's paper: in spite of a long discussion (p.74-77) on the importance of values, Sugden ultimately identifies values with conventions and normative expectations. However it is not clear, as I will discuss below, that values on one hand and norms and conventions on the other are in any way the same thing.

Two attempts that *have* been made to theorise values have been by Tabellini (2008) on one hand and Bisin and Verdier (2001) on the other. In both papers it is assumed that altruistic traits are passed down through a process of "imperfect empathy" from one's parents. Parents are "imperfectly altruistic" in that they judge the best interests of their children using their own, rather than their children's, utility functions. Parents then expend effort in trying to socialise their children with these cultural traits through an expenditure of effort. Children sometimes pick up these traits but at other times pick up a trait from the population as a whole. Children only definitively pick up a trait when their parents and the population as a whole share the same preferences. Tabellini extends this model to account for the fact that some individuals may be more socially or physically distant than others.

As it stands, this analysis is attractive in that it describes much about the transmission of cultural traits within a population. In particular, it neatly describes the problems of parents trying to instil particular values in an environment that may be hostile to such values and also the idea that one tends to value people further away less than those close to. However, there is little in either paper to explain how values emerge within a population. Why should one set of values be seen as something valuable that should be passed on from one generation to another? In other words these papers, while insightful, say little about how values emerge or come to be held extensively within a population in the first place.

A deeper problem with these papers is that it is not clear what a value actually is. It is described as a trait that is inherited to a greater or lesser extent from parents (or role models) to children which in turn gives utility to the parents. However, it is unclear in the models why these traits are valuable to parents or why they think that these traits are valuable to their children. Traits are merely aspects of individuals that are transmitted with greater or lesser effect from one generation to another. They are obviously preferred in some sense to other possible traits but they do not seem to be preferences in themselves. Furthermore, it is assumed that self -interest is separate from values but this is not obviously true- some values may be based on self-interested motivations (one may value being in a trade union because of the benefits it brings to oneself.)

Furthermore, socialisation is something that is very loosely defined. It is obvious that in these models it is a costly process and parents need to make a value judgement in order to socialise their children. It is also an implication of these models that the children take on these values in opposition to their own preferences, which implies that socialisation is to some extent coercive. However, this does not seem to be the case in many circumstances where socialisation, whether from parents or wider society, seems to be voluntary. In this case parents would have little cost of socialisation and people are quite willing to become socialised into another culture and “fit in” with the local population. However, the details of this are left vague by the extant models and need to be examined in more detail.

## 3.Multiculturalism

In this paper the model put forward will be used to give a preliminary, simple analysis of multiculturalism. There are many different views of multiculturalism so, for the purposes of this paper, we will focus on the ideas put forward by one author. The author selected is Parekh (2006) who emphasises the role of culture as a system of beliefs and practices where values play a central role. Everyone, in Parekh’s view, belongs to a cultural community which shares both practices and beliefs. Society and culture are to be distinguished from each other since society consists of a group of humans and the relations between them while culture provides the “software”- the content and principles underlying those relations.

Culture, to a large extent, shapes and influences individuals. In particular, the shared beliefs of people within the cultural community means that people are more likely to follow a given practice. However, culture is malleable so that people within a culture are free to leave the culture (albeit retaining some vestiges of it even after having rejected it.). Cultures, and the individuals within those cultures, are able to take on ideas and values from other cultures. Diversity is inevitable (and, according to Parekh, desirable) within societies and it is quite possible to have cultures living side by side with each other within a society. Even within an initially homogenous society, the diverse capacities of humans within the society mean that there will develop considerable divergences from the cultural community's "norm".

This view of multiculturalism has become quite influential within cultural studies so it is worth asking whether the implicit claims made hold up to scrutiny. Is it indeed possible for cultures to survive in a stable state next to each other in a society? If humans are indeed free to look at and take on board values from other cultural communities will this lead to the cultures surviving or will one set of values simply collapse? Do divergences from cultural norms necessarily remain or will there be a tendency for these divergences to vanish (or even take over the entire cultural community?).

## 4.Grounding for a theory of values

The aim of this paper is to present a theory of values that explains the formation and distribution of values within a society. As such, the models we will create are an attempt to formulate how values may form in a "state of nature" without any institutions[1] and with a population that is free to learn from all other members of the population. They also exclude any influence from education or from parental guidance. This means that members of the population formulate their ideas about values and their actions given these values without any indoctrination or coercive socialisation[2] of values.

By its very nature, these restrictions exclude a large proportion of methods by which people do come to hold their values. As a matter of fact, most people probably do get their values from one of the sources mentioned above. One example is in education. Many people are taught fundamental values while being educated (objectivity in scientific research for example). These values do not necessarily arise from interactions between individuals in a population. Indeed, the values may be formulated by a comparatively small group of people (say, scientists and philosophers throughout history) while the education system is used to spread these values to the wider population.

The model also ignores the effect of role models in a society. Fashion, for example, is an example of a set of values that tends to be set by a comparatively small group of people (actors, models, socialites, pop stars) and is spread through the wider population by the mass media. In medieval times local holy men may have played the same role for religious people.

[1] See Denzau and North (1994) for an example of a theory where institutions *do* influence values.
[2] It will be argued later that "non-coercive" socialisation is little different from ordinary learning

Role models have many more links to other people when it comes to persuading people to change their mind about values. It is assumed in our model that such people do not exist.

The models, therefore, are narrowly defined. However, their aim is to abstract away from the complications outlined above and find out how values can be defined and modelled in a simple context. Once it is seen that values can be modelled with clarity and precision then it will be easier to incorporate the complications.

A crucial assumption made in this paper is that self-interest is not the only motivating force for human beings (while still being *a* motivating force). This should not be a controversial assumption as a multitude of economics and psychological experiments have shown that, even in economically important decisions, self-interest is not always the deciding factor (c.f. Bowles and Gintis 2011, Fehr 1999). As has been pointed out by Mansbridge (1998), it is impossible to seriously base morals on self-interest since it goes against the innate emotions and cognitive capacities of human beings. This does not exclude self-interest since it is possible that in many situations, such as when one is trying to preserve one's life, the utility of one's self-interest would be very high.

The paper will take no position on how humans behave non-selfishly in evolutionary terms. It will be assumed that non- self- interested behaviour has minimal implications for the *fitness*[3] of the individual involved and so fitness will not enter into the utility functions of the individuals making altruistic choices. Individuals will take notice of payoffs involved insofar as these impact on utility but the impact may not be self-interested. One may donate to charity, for example, and gain greater utility from donating more money.

Another problem with the self-interested view is that *one can make an argument for self-interest being a type of value*. A cynic might argue that free-market ideologues do this all the time but there are a vast number of situations where being able to follow your own self-interest is generally valuable. Simply choosing which apple a person would like to eat is a privilege that most people enjoy and find valuable yet it is fundamentally self-interested. Self-interest therefore is not really a special state apart but rather an alternative set of values that a person may select from those that are available.

[3] Indeed it is hard to see why fitness is emphasised by so many game theorists when discussing humans in society. One can carry out many altruistic actions without this in any way reducing one's ability to reproduce. A quick examination of demography, for example, shows that there is no positive correlation between reproductive success and wealth.

## 5.Values as Preferences

First, it is necessary to define values. One definition is by Rokeach (1973):

"To say that a person 'has a value' is to say that he has an enduring belief that a specific mode of conduct or end-of-state existence is personally and socially preferable to alternative modes of conduct or end-states of existence"

It can be seen that this definition is very broad. Indeed one might sensibly ask what the difference is between values on one hand and preferences as defined by economists on the other. It will be argued in this paper that there is no real difference: values are simply a type of preference (although there are preferences, a sudden craving for ice-cream for example, that are not values).

One crucial distinction between values and other types of preference are their comparable longevity. One can say that someone has a value when they hold a definite preference for something for a long time. It would be peculiar to say, for example, that one valued one's wristwatch if one thinks that one will keep it now but was thinking about throwing it away five minutes ago.

One can also say that values are consciously and intentionally held. They are not the result of accidental selection or unconscious urgings. They may be held as habits without being deliberately chosen every time they are invoked but this does not mean that they are unexamined. Even habits usually start out as conscious decisions and are usually the result of intentional behaviour.

A crucial aspect of values (contrary to Sugden's claim mentioned above) is that they are distinct from norms and conventions. A convention is a mode of action that is repeated over a long period of time as the result of an implicit agreement between the parties carrying out the action. This is usually modelled in evolutionary game theory as a particular Nash equilibrium in a game with multiple equilibria. If the population converges on one of the equilibria then that equilibrium becomes the convention. A norm is a convention that evolved previously which is upheld even when there is no history of interaction between the agents who are currently interacting. One uses a normative action because one has legitimate expectations (Sugden 2005) that the other person will also use that norm.

However, a critical aspect of values is that they may hold even if there is no action played corresponding to that value. To take an example: one may hold that stealing kestrel eggs from nests is wrong without ever seeing a kestrel egg that was in a position to be stolen. Likewise, one may value patriotism in wartime without ever having to fight a war. In both cases there is no underlying convention that forces one to follow the moral rule. Likewise one may not have a chance to indulge one's legitimate expectations by following a norm.

It follows from this that values cannot necessarily be accessed by revealed preference. Revealed preference would reveal the existence of conventions and norms but would not say anything about values that are not exemplified by actions. This is reflected in the practical sphere by the use of questionnaires to elicit values rather than revealed preference techniques (c.f. Inglehart et al. 1998).

Values, as defined Rokeach, can cover both individual and group values. While individual values are important to the person concerned, economics is more concerned with the actions of large groups of people rather than the idiosyncrasies of individuals or small groups. As such, this paper will focus on the behaviour of values held by large groups of people rather than individuals.

Another important point is that values seem to be context sensitive (Seligman & Katz 1996). Different values seem to be invoked in different circumstances. When different issues are at stake, different values are invoked to argue them out. A ready wit, for example, is more valued at a dinner party rather than at a funeral wake. This has interesting implications: values are not as universally or consistently applied as one might think and one might apply values differently in different circumstances. It follows that any theory of values must take account of this context sensitivity.

Another point that has to be made about values is that they seem to vary consistently between groups. These may vary between groups in a society (e.g. different food restrictions in the Sikh and Muslim religions in India) or between societies (Inglehart 1998). The latter is particularly important in cultural economics where these differences between countries drive many of the differences in economic variables that have been found in the literature (de Jong 2009). These differences in values have been tested experimentally (see Henrich et al. 2004) and there are distinct differences in reactions between different groups.

The experiments by Henrich et al also demonstrate that any model of values must allow for the fact that different groups may end up with different values in the same situation. It would be unreasonable to postulate, for example, that there is a "long-run equilibrium" set of values to which all people in a given situation will converge. Furthermore, one has to account for the fact that when different groups undertake different actions they seem to give different reasons for these actions as well. It is not just that people converge to different equilibria within the same framework. Often their mental frameworks seem to differ as well.

## 6. Fundamental ideas for a theory of values

Discussion of values is widespread across a variety of disciplines. In this section we will discuss the ideas of two writers- one a philosopher, Joseph Raz (2003), and the other an anthropologist, Dan Sperber (1996). These two writers have similar views on many aspects of values and have tried to accommodate many of the points outlined above, albeit from differing perspectives.

Raz approaches the problem of values from a philosophical perspective. His aim is to try to provide a view of values that allows for value pluralism without degenerating into relativism. In his view, values initially emerge through the use of supporting practices. As people carry out these practices then values are created by the participants in these practices to regulate and excel in them. An example of this is opera which is a practice that has a variety of values (good operatic singing, intelligent interpretation of the music) that could only exist together within opera.

These values, once they are established through the practice, have an independent existence of their own. They can survive even if the original supporting practice dies out. These values tend to be highly specific in that they only apply in the tightly defined areas where they emerge. However, once they emerge then they can be used anywhere within this area without needing a supporting practice. More general values emerge as generalisations of more specific values and do not have their own sustaining practices.

It should be pointed out that Raz is not claiming that sustaining practices "create" values- valuation is a human activity that can only be done by human beings. Humans value things but which things are to be valued is dependent on the social practices in existence. Raz therefore is not claiming that valuation has any objective basis but merely that the occurrence of values is based on contingent practices.

Sperber (1996) approaches culture from an anthropological point of view. In his view the basic building block of culture (which includes values) is the *representation*. The term "representation" has a wide range of meanings but for, the purposes of this paper, the type of representation we (and Sperber) are interested in is the *mental* representation[4]. A mental representation is an interpretation of the world held internally by an information processing device (such as humans!). This is a *cognitive* picture of how the human mind works, relating the mind to information- processing by a computer.

---

[4] A similar notion of representations as the grounding for values is put forward by Mandler (1993- see also Ross and Nisbett 1991), although he tends to refer to a particular type of representation known as a *schema*.

This notion of the mind as being modelled as a computer has its origins in the work of Simon (Simon 1959). Under Simon's ideas, humans are seen as boundedly rational computing machines who construct "specifications" (Simon & March p. 172) of the situations in which they operate. These "specifications", which can be seen to be equivalent to the "representations" used by Sperber and others, consist of knowledge of future events, knowledge of alternatives available for action, knowledge of consequences attached to alternatives and rules or preferences for ordering consequences. This defines a situation for a particular actor but does not give an objective assessment. Indeed, this assessment is usually a screened, simplified and biased view of the situation. Knowledge of the future, for example, is quite often simply unknown and is usually dealt with by using previously used ideas.

This idea of representations is used widely within Social Psychology. According to Ross and Nisbett (1991 –see also Bowles & Gintis 2011 p.43) situational variables have a substantial effect on human decision making and quite often make more of a contribution towards decision making than the character and stable preferences of the decision maker. Furthermore, the representations of situations are constructed by the decision maker and so are essentially subjective rather than objective. They can vary from person to person and can vary within a person from one time to another. Indeed, it has been argued by Sen (1980) that description of any kind is inevitably selective and partial so insofar as a representation is an individual's description of the external world, it will suffer from the same flaws for the same reasons.

The lack of knowledge of the situation and the arbitrariness involved is particularly marked when it comes to values because of the lack of any objective link between values and a particular objective situation (See Bowles & Gintis 2011 p. 44 for a similar claim). As Raz points out, valuation is essentially a human activity and any attempt to derive values from objective reality is probably futile. It follows that any representation of a situation when a person is valuing that situation will not have totally objective foundations. No situation will logically suggest any particular set of values independently of human valuation. Representations therefore have an inescapably subjective element when referring to values.

Mental representations, according to Sperber, are used by the information processing devices (i.e. minds) to affect other people by interfacing with the physical environment. This can be by direct or indirect communication and this in turn causes other people to create mental representations in their own minds. As a result of this, mental representations are transmitted from one person to another. Sperber then defines a *cultural or social representation* as a representation that is commonly held among the general population. He sees the aim of anthropology as being to investigate the “epidemiology” of these cultural representations.

One question that Sperber attempts to answer is which sort of mental representation will succeed in becoming a social representation? This is a wide ranging problem but some relevant attributes are:

i) Ease of memorisation

ii) Existence of relevant background knowledge.

iii) A motive for communicating the content of the representation

iv) Recurrence of the situation which the representation gives rise to.

v) Availability of external memory stores (writing, mobile phones)

vi) Existence of institutions engaged in the transmission of representations.

This will tend to cut down the number of mental representations that can ever become social representations. One side effect of this is that many representations that are deemed to be important in society- knowledge of the law or of science for example, rarely become common knowledge among the mass of the population because of their complexity, lack of knowledge etc.

## 7.Values and Emotions

Mandler (1993) emphasises the crucial role played by values in the inducing of emotions. Mandler points out that, however visceral and “hot” an emotion may be, all emotions must rely on some cognitive processing so that a person may establish and interpret a situation. A person usually only gets angry at having their wallet stolen if they have, indeed, had their wallet stolen. A person who gets angry at having their wallet stolen when this is not the case is seen as having no justification for their anger. To get angry, therefore, requires cognitive processing to establish that the wallet has indeed been stolen.

Many, but not all, emotions tend to be aroused by an expectation generated by a mental representation. This of course does not apply to all emotions. Many emotions are instinctual and automatic so that there is little beyond identifying the situation that the mind does in sparking the emotion. If an iron rod is being dropped from the roof of a building then, once this has been identified, the body tries to avoid it, automatically invoking a feeling of fear. However, many such emotions are learned and are acquired over time and it is these emotions that often emerge from following, or failing to follow, values. People, therefore, will identify the situation and also the value that needs to be followed before emotion emerges in response. Given that values are assumed to be preferences expressed within a certain representation, it follows that many people will feel emotional if values are being violated or even if values are being followed.

Many emotions therefore have their foundations in values or in the violation of perceived values. Frank (1988) has observed that emotions can override self-interest and that this enables people to make commitments which solve problems that cannot otherwise be solved[5]. However, the ability to make commitments means that it must be possible to reliably communicate one's emotions. Frank claims that this is achieved either through a person building a reputation for honestly having these emotions or by outward signs of this emotion that are very hard to fake[6].

It should be noticed that Frank is not claiming that these emotions are on public view or that they cannot be faked. However, Frank points out that that, like all signalling devices, emotions are quite hard to fake and also that, for emotions to work as a commitment device, most emotions must be genuine. While it may be difficult to find out another's true emotions, it is rarely impossible and fakery is often unsuccessful. Indeed, people who tend to be successful at faking their emotions (such as psychopaths) tend to be ostracised within society and quite often end up worse off. Indeed, from an evolutionary point of view, it is difficult to see exactly why emotions would survive if they were easily faked. Their role as a signal would be undermined by rampant fraud, while being psychologically costly to those experiencing the emotions.

It follows that we have a linkage between values and emotion with the causality going from the values to the emotions rather than vice versa. The existence of values induces certain kinds of emotion when they are followed or not. Emotions are very hard to fake and are also observable by other individuals. Since they are hard to fake then they act as a reasonably reliable signal to the other individuals as to the underlying values being expressed.

[5] An example of this would be the emotion of love in marriage which would prevent a person from bailing out at the earliest possible favourable opportunity. A person who can show that they love another person would be seen as a better marriage partner because they are in the marriage for the long term and so they can invest more in the marriage.

[6] See Scharlemann, Eckel et al. (2001) for an economic experiment that demonstrates that participants react positively to photographs of their opponents smiling in a trust game.

## 8. Building blocks of a theory of value

The discussion above has shown that there is room for a theory of value which incorporates the notion of representation from cognitive science. This paper will follow through this idea in incorporating it within a game-theoretic model. However, this creates problems because there is no accepted way in which representations can be modelled within game theory. The usual method of looking at heterogeneity among agents is to use the concept of type. However, this is obviously insufficient since types are assumed to be fixed for each person and out of the conscious control of the player (Harsanyi 1967) while one can always change one's representation. While some models do try to model psychological issues using the types framework (e.g. Benabou & Tirole 2006) they suffer from the problem that the psychology for each person is unchangeable and there is no space for changing one's mind.

For the purposes of this paper we will use a model called a *multirepresentational* game (see Jones 2016 for more discussion on this). This is defined as a game that allows two or more representations within the game structure for one or more players. Games in the economics literature usually only allow one representation that covers all players in the whole game. The multirepresentational game differs in that an action in a game can allow a representation to change partway through. In particular, we will focus on games where each player has two possible representations. Since the choice of representation precedes the choice of physical action, these games are necessarily extensive form games. The representation that is currently chosen is known as the *active* representation.

It is claimed here that the process of picking a representation is a conscious, intentional process by the agent. As such it is a species of *mental action*. Mental actions are processes in the mind that have as their goal another mental process (Proust 2001,2009). Examples of mental actions are easy to come across. One may try to remember when the bus is going to leave. A purchaser may try to work out what the VAT is on a purchase of window frames. A student tries to concentrate on a lecture. In each case an agent is deliberately trying to engage some mental faculty.

This notion of mental action is quite old and goes back to the Enlightenment. Locke (1689/1998) believed that the mind deliberating and operating on itself and its own ideas was the bedrock of philosophy. Geach (1957) presents a more recent 20th Century view of mental acts, particularly in relation to judgment. Mental actions are much the same as ordinary, physical actions in that they have a goal, are intentional and are not the result of unconscious thought. As intentional actions they can be rationalised by the beliefs and desires attached to the action- in other words the beliefs and desires motivate the action and act as a causal explanation for that action (Davidson 1963). Mental actions therefore can act as the means by which representations enter into games since the action can determine how a person actually thinks and the reasons given constitute the representation.

It follows that one can model the picking of one representation out of several[7] as a simple choice between actions as one would do with ordinary physical actions. Essentially it is assumed that the picking process is the result of individuals judging whether a representation is suitable for a given situation. The "suitability" is represented by the expected utility of that representation. This way of modelling representations is not unusual as Henrich and Boyd's (2002) first model is essentially a model of this process within a population. However, the model presented here puts this explicitly within a game- theoretic context.

In order to motivate the following discussion we shall introduce the notion of a *base game*. A base game is a normal form game structure G=⟨P,S,Q⟩ where P is the set of n players, S is an n-tuple of pure strategy sets (one for each player) and Q is a set of outcomes. A base game, it should be noted, has no explicitly defined utility payoffs but only outcomes specifying the event that occurs. An outcome may have some material payoffs or it may have something which happens that a players likes or approves. However, a base game does not include any preference structure. Instead the base game illustrates the number of players and the strategies given to each player, while leaving the preference structure open. In this paper we will assume that we are dealing with two player- two strategy base games.

The purpose of a representation is to impose a set of expected utilities[8] onto a base game. Each representation imposes a different set of expected utilities according to the interpretation of the situation and the values being considered. As a result of these different interpretations, greater or lesser preference will be given to one outcome in a game matrix over another. Representations therefore will reflect the context-sensitive nature of values in terms of utilities that vary according to the situation and the representation of that situation. We will assume, following Sperber that these representations are *social* representations.

Strictly, it will be assumed that the utilities assigned by representations will be *expected* utilities comprising probabilities and utilities. The utilities represent the valuations

[7] Note that this notion of choice of representations is similar to a choice in descriptions as outlined by Sen (1980)

[8] The expected utilities here are assumed to be conventional von- Neumann- Morgenstern expected utilities.

prescribed by the representation. The probabilities represent the belief that certain things may happen or are the case. This encapsulates the idea that culture and values are partially determined by the beliefs of the cultural community. These beliefs may be associated with, say, religion or maybe the importance of certain historical events. A Scotsman may, for example, value wearing a kilt because he has a belief that it is an item of clothing that has a history over a thousand years old[9] and also attaches a high value, reflected in his utility, to that historical background. These probabilities are *not* connected to the proportions of people with particular representations in the population.

In Jones (2016) representations are operationalised further in terms of attributes and predicates based on those attributes. This allows a situation to be more exactly described by the subjective attributes assigned by individual players. Meanwhile, the predicates allow representations to be explicitly described, as in the cognitive science literature (see Fodor 1989) as *propositional attitudes* i.e. attitudes expressed as propositions. Utilities are then built up over relationships between these predicate- based representations. While this view of representations is useful, it is not essential for this paper and so the attributes and predicates will be supressed. Utilities will therefore be represented as constant values.

A 2x2 base game can therefore be transformed into an ordinary 2x2 game by the use of a representation. Assume that this is an evolutionary symmetric game being played by one population of players. In that case the payoff matrix will be as follows:

| | Player 2 | | |
|---|---|---|---|
| Player 1 | | Up | Down |
| | Up | a,a | b,c |
| | Down | c,b | d,d |

Where a,b,c,d are all utility payoffs. In that case, according to Weibull (1995) there are four categories of game:

Category 1:a>c; b>d; this leads to a Nash Equilibrium at (Up, Up)

Category 2: a>c; b<d: this leads to two Nash Equilibria at (Up,Up) and (Down, Down)

[9] This is a false belief- see Trevor-Roper (1983). Beliefs need not be true to be included in a representation. All such probabilities of belief are assumed to *subjective*.

Category 3: a<c; b>d: this leads to a mixed Nash Equilibrium at $(\frac{b-d}{(c-a)+(b-d)}, \frac{b-d}{(c-a)+(b-d)})$.

Category 4:a<c; b<d; this leads to a Nash Equilibrium at (Down, Down)

In this paper it will be claimed that the preferences under given representations attached to the base games are to be identified with values.

**Definition**: A value is a preference relation between outcomes within a particular representation.

We will focus here on changes in choices as a result of changes in numbers in a population holding a particular representation and the utilities attached to that representation. As Ross and Nisbett (1991) point out, other people are generally one's best source of information on the world around us and so the initial spread of representations in the population will assumed to be by a process of diffusion of social representations. It has been shown that most people dislike being in tension with the group of people around them and will not only vary their behaviour but also their *understanding* of a situation in order to fit in. This suggests that not only will one's strategy choice in the base game change in line with that of the group and one's utilities but also one's choice of representation as well.

There are various issues that may seem to be problematic for this model. The first is how an individual is supposed to make a choice within the game as a whole. The difficulty is that a representation changes the utilities of a given outcome. This means that there is a potential contradiction between the utility of an outcome when one plays it and the utility as one looks at the game as a whole. In the former case, one's utilities are formed given that one *has* adopted this representation. In other words one's utilities are imagined indicatively. However, if one is assessing the game as a whole then one is forced to evaluate a situation if one *would* adopt this representation. This means that one's utilities are imagined subjunctively. These two conditional utilities are not necessarily the same and this would mean that the game as a whole is not consistent. Jones (2016) analyses the conditions under which these utilities do coincide. It will be assumed here that the two sets of utilities do coincide and that the game is consistent.

A related issue is how the representations are judged. It is assumed that the utility of the representations is summed together with the utility of the outcomes. However, there is the issue of how the representations are judged against each other, given that representations are supposed to represent a player's thinking about utilities in the game. It will be assumed that the utility of each representation includes a “common core” that allows comparisons to be made between the representations. All representational utilities therefore are inherently subjunctive in nature (see Jones 2016) and can be compared across representations.

As has been argued above, this paper rejects the idea that agents are *necessarily* self-interested. Instead, self-interest is one possible attitude that one could incorporate into a representation and would compete on an equal basis with other possible attitudes. This is what allows the framework put forward here to work. If self-interest is imposed as necessary then the scope for differences between representations would be cut down to those that are monotonic with the material outcomes- which would be a very meagre offering. In fact no distinction is made between self-interest and other motivations in that there is no separation between self- interested preferences and non-self-interested preferences. All are subsumed under generic preferences which may be self –interested or otherwise. Whether a person behaves in a self-interested manner depends on the representation.

Given that a choice is made between representations then the question has to be asked as to why representations cannot be changed simply to suit the convenience of the agent. If one ends up in an unfavourable situation as a result of adopting a representation then why can one not simply rewrite one’s representation? The choice of representation is part of an extended game where one is brought into equilibrium in strategies in the base game but also in representations. Representations are therefore fixed by being in equilibrium and one loses utility by deviating from it. Rewriting a representation involves creating a new game with a larger set of representations that has its own equilibrium which also fixes the representations. Also the model allows for other attitudes apart from self-interest- among them an interest in accuracy and objectivity which would pull against self-interested motives. One cannot assume that the agent *will* choose a self-interested representation.

One issue is that representations, operating as mental actions, are not visible to the other player so the other player is not aware which representation has been chosen. We will assume throughout the paper that all available representations are known to all players. However we will construct two models to discuss two different approaches to this issue. One is simply to treat mental actions as hidden actions and to model accordingly. This will be done in the second model

The second possibility, following the discussion in section 7 is that values evoke emotions and that these emotions can, fallibly, be observed by other people. It follows that we will make the same assumption in the first model and assume that the other player’s chosen representation will be available to the first player. While it is acknowledged, following Frank (1988) that this is costly, this costliness will not have any effect on the model since the costs will be the same whichever representation is chosen. This will be done in the first model.

Even though the models have not been fully explained, one can see that this method has the advantage of satisfying some of the criteria for a theory of value. First of all, not all values can be detected by a researcher through revealed preference. In a category 2 representation, the population may converge to playing (Down,Down) but this simply reveals that d>b. It does not reveal that a>c. The latter would have to be discovered by a researcher by other means. This fits in with the idea that values do not always correspond to actual choices.

Secondly, the use of representations is context sensitive. Different contexts evoke different representations and these in turn evoke different utilities as payoffs in the game. The use of representations therefore explains why a given base game with the same actions (e.g. whether or not to tell a joke) may result in different actions being chosen in different contexts (a dinner party versus a funeral). Representations also explain why different cultures choose different actions in similar choice situations. This is because different cultures provide different reasons and hence different representations for choosing one action over another.

The model explains the longevity of values since equilibrium in the model fixes the representation and this in turn fixes the preferences in the model, whether they are actually revealed or not. It also explains why large groups of people may come to hold certain values rather than others- there is an equilibrium in representations which determines the values of these groups. Also, in many respects, insofar as the model has multiple equilibria, it explains why different societies have different values- the populations converge on different equilibria.

Finally, this gives us some insight into the idea of "socialisation" that is often employed in the literature. Often socialisation is assumed to be a process that goes against the self-interested preferences of the recipient of the process. However, it is unclear why it should happen at all- why should someone accept a process that will end up with them choosing a less- preferred outcome? This is solved within this model by (mostly) eliminating the assumption of self-interest. Self-interest in representations competes equally in utility with other attitudes and one chooses the attitude with the highest utility in the representation.

## 9. The model

The preceding discussion leads on to a very simple modelling strategy for values. Every person who plays a game first of all chooses a representation and, given the utilities attached to the base game, then chooses an action to play in the base game. We will refer to the choice of (behavioural) strategy at the representational level as a choice of representation while the choice of behavioural strategy in the modified base game will be called the *physical strategy*. The role of socialisation will be approximated by the use of evolutionary mechanisms applied to extensive form multirepresentational games. As mentioned before, two models will be constructed.

The first model has the structure of a sequential, two-player, symmetric, evolutionary game with one population and no moves by nature. More precisely, it is a two stage- two strategy, simultaneity game where payoffs are cumulative between stages (Cressman 2003 p. 192). This means that, when playing a game, the two players simultaneously choose their representation and then, once the representations have been chosen, they choose which action to play given the utilities established by the representation. This leads to an extensive form game with four subgames, each in the form of a 2x2 normal form game. It is assumed that the player knows which combination of representations is held by both players before they choose their actions.

Cressman (2003) has put forward a method by which extensive form games can be modelled by evolutionary methods while still taking into account the sequential nature of actions within the game. Cressman demonstrates that one can decompose a replicator dynamic for the normal form of the whole game into separate replicator dynamics for the subgames and the truncated game. This can be done by assuming that the replicator operates on the Wright Manifold- the combination of probability points where the choice of strategy in one subgame is independent of the choice of the same strategy in another subgame. The game as a whole can be brought into equilibrium by first of all finding the equilibria of the subgames and then substituting one from each into the truncated game. It can be shown that when the subgame and truncated game replicators converge on an equilibrium point then this is a subgame perfect equilibrium for the whole game.

The game as a whole is played by one population. In normal form this is fairly straightforward as this just requires a one-population set of equations for the strategies played in the game. When the game is decomposed according to the Wright manifold then, for the truncated game and any symmetric subgame, the situation is still the same. Each one is treated as a smaller, one-population game. In the case of a two-stage, two strategy simultaneity game this would mean that each symmetric subgame and the overall truncated game would have a one-equation replicator dynamic equation.

However in such symmetric extensive form games, there will generically be asymmetric subgames as well as symmetric subgames. In this particular type of game there will be two such games. This means that it is not appropriate to treat these games as symmetric. However, it is also not appropriate to treat the two sides of the game as members of different populations as it is assumed that the whole game is played with one population. In this case, Cressman recommends treating the two games as components of a single linked asymmetric game where each "player" in the game is actually a role selected within the game by a player from a unified population. One set of choices leads to the player playing one side of the game while another set of choices leads to them playing the other side of the game.

The replicator dynamics used in this model are used as an approximation to social learning rather than having a biological context. As such one can see members of a population as having a tendency to play a given strategy, the greater the difference in expected utilities between one strategy and the average expected utility of all strategies. The biological interpretations of the replicator dynamics involving concepts such as fitness, reproduction and multiple biological generations are ignored here (see Binmore 1988 for one possible interpretation along these lines).

Each game played by the population has the structure in figure 1. $R_1$ and $R_2$ represent the representational choices made by player 1 and player 2. The information set covering player 2's nodes implies that each person chooses their information set simultaneously, although the nature of their choices is known to each other later on in the game. The subgames starting from $u_1$, $u_2$, $u_3$, $u_4$ represent the different interpreted versions of the base game. The subgames emerging from $u_1$ and $u_4$ represent the situation where both players have chosen the same representation. The asymmetric subgames following from $u_2$ and $u_3$ represent situations where the two players have chosen different representations.

{Figure 1 Here}

The strategies in the game correspond to the two types of action discussed above. $R_1$ and $R_2$ correspond to the mental actions of choosing representation 1 and representation 2. U and D are physical actions corresponding to moves in the base game. These are the moves that actually result in a material outcome being achieved.

It is assumed that any utility deriving purely from the choice of representations is incorporated into the final utilities in the terminal nodes. It is also assumed that, if a player chooses a certain representation and physical action (given their opponent's physical actions) then they will have the same payoff irrespective of the representation chosen by their opponent. This means that a player's final payoffs are not influenced by the representation chosen by one's opponent.

The use of this model allows us to model many of the basic facets of the discussion above. The choice of representation is a mental action where individuals deliberately choose which representation to use in analysing the situation. The extensive form structure of the game allows this mental action to be taken prior to the physical action and represents a settling of the mind of the player on one particular interpretation. The simultaneous choice of representations by agents models the fact that agents do not know each other's choices before they make their own.

It is assumed that players know each other's representations before they play their physical actions. This reflects the previous discussion of Frank's and Mandel's work where it was assumed that one's view of the world would be accessible to the other player as a result of the visibility of their value-induced emotions. Meanwhile, the evolutionary structure allows one to explain the intuition common to Raz and Sperber that values are essentially social phenomena that emerge through social interaction in the population.

More specifically, the model also catches Raz's intuition that the occurrence of values depend on social practices. If one sees the physical games being played as social practices then the representations that are selected out by the evolutionary processes can be seen to depend for their existence on those practices.

Note that the game allows for two representations to be chosen which lead to four subgames (with roots at $u_1$,$u_2$,$u_3$ and $u_4$) in figure 1. The subgames following from $u_1$ and $u_4$ will be referred to as *symmetric* subgames and correspond with the following normal form matrices:

| | Player 2 | | |
|---|---|---|---|
| Player 1 | | U | D |
| | U | a,a | b,c |
| | D | c,b | d,d |

and

| | Player 2 | | |
|---|---|---|---|
| Player 1 | | U | D |
| | U | w,w | x,y |
| | D | y,x | z,z |

It follows that there are two asymmetric subgames following from $u_2$ and $u_3$. Given the use of Cressman's model these can be seen as two different roles in one conventional asymmetric game as shown in the matrix below:

| | Player 2 (From $u_3$) | | |
|---|---|---|---|
| Player 1 (From $u_2$) | | U | D |
| | U | a,w | b,y |
| | D | c,x | d,z |

This means that, when a player reaches $u_2$, then they will take the role of player 1 in a two-population asymmetric evolutionary game while if a player reaches $u_3$ then they take the role of player 2 in that game.

The structure of the games means that the subgames act as representations of the original base game. Nothing is said within the model as to the relationship between any material payoffs in the base game and the representational subgames within the game in figure 1. The model does not deny that some links may exist but simply does not make them explicit. This is completely conventional- strictly, within Von Neumann and Morgenstern's (1953) game theory, *all* payoffs are in terms of utilities and none of the material payoffs are specified. In addition to this, the conventional theory of utility makes no assumptions about self-interest and neither does the current theory. This means that the scope of the theory is very broad.

For the purposes of this paper the representations consist of symmetric games using the four categories outlined in section 8. It is assumed that even if the two representations in each game are in the same category (i.e. if the utilities in each representation have the same ordering) they still have different levels of utility.  This gives some idea as to what happens when one has similar but not identical representations of a situation or when different parts of a population have different levels of "enthusiasm" for a similar view of a situation.

The term (XvY) will be used to denote when a game involves a representation formed from category X and a representation formed from category Y. So, for example, (1v2) means that the game has two representations from categories 1 and 2. The representation from category 1 would be represented by utilities a, b, c,d while the representation from category 2 would be represented by utilities w,x,y,z. Given the overall symmetry of the game, it can be seen that there is no difference if the two categories are interchanged.

The second model is different from the first model in that it is assumed that the representation picked by the other player is hidden. This causes significant changes to the model because, when one plays a given physical strategy, one cannot tell whether one's opponent is playing representation R1 or R2. It follows that a physical strategy will have four potential outcomes (depending on whether the opponent plays {R1,U}, {R1,D}, {R2,U} or {R2,D}). Although the player will not distinguish between whether an opponent's physical strategy has been played under R1 or R2 because the payoffs are the same, this framework still has a considerable influence on modelling.

One issue is how to model the fact that each person in a game picks their own representation but this is unknown to the other person. This is solved by having each person having a 50% probability of playing one of two roles in the game. If they are selected by nature as role 1 then they act as player 1, selecting a representation while not knowing player 2's representation. Having selected his own representation, Player 1 then plays a 2x4 game with player 2 with the four strategies being the representation/physical strategy combinations mentioned above.

If a person is selected by nature into role 2 then they act as player 2. Player 2 then selects a representation while not knowing player 2's representation. Then, as previously, player 2 plays a 2x4 game with player 1 with player 1's choice of representation being hidden. This leads to a game that has four subgames, each of which is a 2x4 game This structure of game can be seen in figure 2:

{Figure 2 here}

It should be noted that figure 2 is a *truly symmetric game* (see Cressman 2003) where the payoffs are identical on each side with only the players swapping round. It follows that the game can be analysed simply by looking at one role and analysing the equations for the evolution of that side. The players act as opponents to each other in the asymmetric 2x4 games but, since the games are the same on each side, the analysis does not need to be replicated.

Much of the description of the game in figure 2 is similar to that in the first model. Each 2x4 game represents the "fusing" of a symmetric game in model 1 with its corresponding asymmetric game. The labelling of the payoffs remains the same and the notation given above for category combinations will be continued here. Again the entire game will be played by one population but the equilibria are more interdependent in terms of variables than for the first model. All the subgames are asymmetric although the players on each side of the game are selected from the same population in different roles.

## 10. Layout of models

Given model 1 above, the extensive form game shown in figure 1 can be modelled as a one population evolutionary game using the replicator dynamic. First of all, examine the normal form of the game outlined above. This game is one with eight strategies (see appendix). The proportions of the population playing a given strategy $s_i$ is $q_i$. p represents the mixed strategy played by one's opponent, while the payoff from playing a strategy $s_i$ is represented by $\pi$ ($s_i$/p). The equation of motion for playing $s_1$ is given by:

$$\dot{q_1} = q_1\left[(\pi(s_1/p) - \left[\sum_{i=1}^{8} q_i\big(\pi(s_i/p)\big)\right]\right]$$

Similar equations can be constructed for $q_2$, $q_3$ etc. These equations, as is shown in Cressman (2003 Ch.6 appendix), can be decomposed into one equation for the truncated game and four equations for the subgames. Suppose that $P_{R1}$ represents the probability that $R_1$ is chosen in the truncated game while $q_U^{u_x}$ is the probability that U is chosen in the subgame following $u_x$ (x=1,2,3,4).

The equation for the truncated game is:

Equation 1

$$\dot{P_{R1}} = P_{R1}(1 - P_{R1})\big((P_{R1}\bar{n}_1 + (1 - P_{R1})\bar{n}_2) - (P_{R1}\bar{n}_3 + (1 - P_{R1})\bar{n}_4)\big)$$

Here, $\bar{n}_x$ are the equilibrium expected values of the subgames following from $u_x$ .

The four deterministic equations for the subgames are as follows:

Equation 2

$$\dot{q}_U^{u_1} = P_{R1}q_U^{u_1}(1 - q_U^{u_1})\left[q_U^{u_1}(a - c) + \left(1 - q_U^{u_1}\right)(b - d)\right]$$

Equation 3

$$\dot{q}_U^{u_2} = (1 - P_{R1})q_U^{u_2}(1 - q_U^{u_2})\left[q_U^{u_3}(a - c) + \left(1 - q_U^{u_3}\right)(b - d)\right]$$

Equation 4

$$\dot{q}_U^{u_3} = P_{R1}q_U^{u_3}(1 - q_U^{u_3})\left[q_U^{u_2}(w - y) + \left(1 - q_U^{u_2}\right)(x - z)\right]$$

Equation 5

$$\dot{q}_U^{u_4} = (1 - P_{R1})q_U^{u_4}(1 - q_U^{u_4})\left[q_U^{u_4}(w - y) + \left(1 - q_U^{u_4}\right)(x - z)\right]$$

It will be noted that all of the equations have the form of replicator dynamics up to the multiplying of the formula by either $P_{R1}$ or ($1-P_{R1}$) which has no effect on the direction of the trajectories. Note that equations (3) and (4) are interlinked which reflects their status as describing two roles in one interlinked asymmetric game.

One issue that must be discussed before analysing the model is the issue of stability within the model as a whole. In order for equilibrium to emerge for the model as a whole, Cressman (p. 193-4 Theorem 7.2.1 and footnote 4) points out that there cannot be cycles within the subgames. This is not a problem for the symmetric subgames where the decomposed replicator dynamics act as if within one population games and so all Nash equilibria are strict. However, within the asymmetric subgames this is a problem since they act like a linked two population game where cycles are possible.

The main problem is the so- called "Buyer Seller" game which is a non-degenerate, two- strategy bimatrix game (see Cressman p.78). In this case the replicator dynamics trace out cycles and do not converge towards the unique mixed Nash equilibrium. An examination of the (2v3) and (3v2) category combinations applied to the asymmetric subgames shows that the corresponding linked asymmetric game is identical to the Buyer-Seller game. Since no other category combinations in the linked asymmetric subgames are identical to the Buyer-Seller game and the Buyer-Seller game is the only one that causes these cycles it follows that the (2v3) and (3v2) category combinations are the only ones that cause this problem.

It follows, therefore, that the (2v3) and (3v2) category combinations should be excluded from any analysis of the model if one wishes to achieve a stable equilibrium. However, this makes perfect sense within the literature, especially in Raz's (2003) theory of values. For Raz, values emerge because of the existence of a sustaining practice. If we interpret equilibrium physical strategies in subgames as practices and the values are incorporated in the representation then we can analyse his ideas in terms of our model. For something to be a practice it follows that it must be stable. Given that, under this model, a strategy played in a subgame is a practice then it can only be stable if it is in a stable, non-cyclical equilibrium. It follows that (2v3) and (3v2) must be excluded on theoretical grounds as well as mathematical grounds.

Another theoretical implication of the exclusion of the (2v3) and (3v2) category combinations is that there are no stable mixed equilibria in the asymmetric subgames. This is because the asymmetric subgames are parts of an asymmetric linked game where the two roles in the game act in the same way as two different populations. In such a case (2v3) and (3v2) are the only combinations that result in a unique interior equilibrium. In all other cases the interior equilibrium either doesn't exist or is not stable.

In the case of model 2, the decomposition is from a 4x4 normal form game in which both players have the strategies {R1,U}, {R1,D}, {R2,U} and {R2,D}:

| | | Player 2 | | | |
|---|---|---|---|---|---|
| | | R1 U | R1 D | R2 U | R2 D |
| Player 1 | R1 U | a,a | b,c | a,w | b,y |
| | R1 D | c,b | d,d | c,x | d,z |
| | R2 U | w,x | x,c | w,w | x,y |
| | R2 D | y,b | z,d | y,x | z,z |

We will assume that the proportions of the four strategies in the population are $t_1$, $t_2$, $t_3$ and $t_4$ respectively. We will rearrange the equations of motion of these four probabilities into three differential equations in terms of three variables: p, q and r. p is the proportion of the population that chooses R1, q is the proportion of the population that chooses U given that R1 has been chosen and r is the proportion of the population that chooses U given that R2 has been chosen. The derivations of the following equations are given in the appendix. Using the following abbreviations:

$$X_1 = (t_1 + t_3)a + (t_2 + t_4)b = (pq + (1-p)r)a + \big(p(1-q) + (1-p)(1-r)\big)b$$
$$X_2 = (t_1 + t_3)c + (t_2 + t_4)d = (pq + (1-p)r)c + \big(p(1-q) + (1-p)(1-r)\big)d$$
$$X_3 = (t_1 + t_3)w + (t_2 + t_4)x = (pq + (1-p)r)w + \big(p(1-q) + (1-p)(1-r)\big)x$$
$$X_4 = (t_1 + t_3)y + (t_2 + t_4)z = (pq + (1-p)r)y + \big(p(1-q) + (1-p)(1-r)\big)z$$

We can state the equations of motion for $t_i$:

$$\dot{t}_1 = t_1(X_1 - (t_1X_1 + t_2X_2 + t_3X_3 + t_4X_4))$$
$$\dot{t}_2 = t_2(X_2 - (t_1X_1 + t_2X_2 + t_3X_3 + t_4X_4))$$
$$\dot{t}_3 = t_3(X_3 - (t_1X_1 + t_2X_2 + t_3X_3 + t_4X_4))$$
$$\dot{t}_4 = t_4(X_4 - (t_1X_1 + t_2X_2 + t_3X_3 + t_4X_4))$$

One can derive the equation of motion for p:

$$\dot{p} = p(1-p)((qX_1 + (1-q)X_2)) - (rX_3 + (1-r)X_4))$$

We can also derive the following equations of motion for q and r:

$$\dot{q} = q(1-q)\Big(\big(qp + r(1-p)\big)(a-c) + \big((1-q)p + (1-r)(1-p)\big)(b-d)\Big)$$

$$\dot{r} = r(1-r)(\big(qp + r(1-p)\big)(w-y) + \big((1-q)p + (1-r)(1-p)\big)(x-z))$$

These three equations of motion comprise the system for the second model as set out in figure 3. It can be seen that the player of the game has a choice in each subgame between two strategies. Each of these strategies has four potential outcomes depending on which of four strategies one's opponent plays. However, in the subgames, the player cannot differentiate between whether his opponent has played {R1, U} or {R2,U}because the payoffs are the same and it is assumed that he does not otherwise know whether R1 or R2 has been played. Likewise he cannot distinguish between playing {R1,D} or {R2,D} for the same re ason.

It follows that the probabilities in each strategy: (qp+r(1-p)) and ((1-q)p+(1-r)(1-p)), represent the probabilities of playing U and D respectively. Players therefore do not directly perceive the probabilities r and q but simply the proportions of U and D in the entire population, which is information they would expect to have. A similar analysis can be made of the equation of motion for p when one realises that the individual only analyses the $X_i$ expressions since the equation of motion for p represents how the proportion of p in the population is updated. Xi contains the same probabilities for playing U and D. It follows that the knowledge requirements for players are remarkably low- they only need to know the proportions of U and D in the population and not those of p, q or r.

This model is not as tractable as that of model 1. The main reason is that the mixed strategy restpoints for the equations of motion for q and r are not singular. This, together with the fact that p, q and r have significant influences on each others' equations of motion means that characterising mixed strategies becomes very complex. Nevertheless, as will be demonstrated in the next section, one can still derive some powerful results from this model.

11. Specific properties of the models

The models outlined above are designed to explain the incidence of particular values within a population. Each representation encompasses a different set of utility numbers to be attached to the base game. Each utility number is a value in that it shows, given a particular representation, how much one outcome is valued. Another representation would give another set of values in a game. It follows that the models given here show how different sets of values may compete and survive against each other.

As such, we would expect these models to tell us something about culture in general insofar as culture consists of value judgements. How far can different value- systems coexist within a society when the society satisfies the models' assumptions? Will one value system be wiped out by the dominance of the other one or is there a natural tendency for value systems to survive quite happily together? Such questions have obvious relevance for multicultural societies where different values may exist within the same society.

This leads on to particular questions that can be asked within the models. Naturally, these models are limited in that they only focus on two possible representations and two possible physical strategies. Any real world system could have large numbers of both. However, the model does allow us to think clearly about how physical actions (i.e. the observable actions) interact with representations (i.e. people's values).

One point that must be made straightaway is that, within the first model, an overwhelming majority of the possible permutations of representations end up with the entire population holding one representation. This can easily be seen by examining the properties of the truncated game. The truncated game is essentially a one-population, two- strategy evolutionary game where the strategies are mental actions selecting representations. In most cases we end up with the four possibilities outlined by Weibull (1995) for generic games which are the same as the four categories used for subgames[10]. For three possibilities the population will end up with just one representation. It is only for the mixed equilibrium (Category 3 applied to the truncated game) that we will end up with more than one representation within a population. It follows that it is rare to have a situation where more than one set of values exists in equilibrium within a population.

[10] It is possible (such as (1v1) or (4v4)) to end up with repeated payoffs in the truncated game. However, this simply restricts the equilibrium possibilities only to pure strategies and will result in convergence to a unique representation.

For the second model this precise conclusion is hard to replicate because of the existence of non-singular mixed strategy rest points. However, one can state the following proposition:

**Proposition 1:**

a) Most equilibria in category combinations containing just categories 1, 2 and 4 are situations where just one representation is being played by the entire population.

b) All category combinations containing category 3 and one of categories 1,2 and 4 have an equilibrium where just one representation is played by the entire population.

Proof: see appendix1

This has the same effect as the discussion above, in that categories 1,2 and 4 are seen as those that only allow pure strategies equilibria in representations.

Another point is that the vast bulk of the possible permutations of representations also end up with the entire population playing one physical strategy. This can be seen by realising that for the population to end up playing different physical strategies involves either mixed strategies in the truncated game or in the subgames (or both). Since only one case in the four possibilities outlined by Weibull (1995) allows mixed strategies (together with the case outlined in proposition 2 below) whether in the truncated game or the subgames then it follows that the majority of the time the population will follow just one strategy.

Again, this result cannot be neatly replicated in the second model for similar reasons to those discussed above. However, the following proposition has much the same effect:

**Proposition 2:**

a) Most equilibria in category combinations containing just categories 1, 2 and 4 are situations where just one physical strategy is being played by the entire population.

b) All category combinations containing category 3 and one of categories 1,2 and 4 have an equilibrium where just one physical strategy is played by the entire population.

Proof: see appendix 1

It follows that, in many cases, there will be little evidence of differences between different parts of the population. However, this does not mean that there will be no interesting cases. One question that is of interest is whether it is possible to have a stable situation where one part of the population has a set of values that results in them playing a given physical strategy while the other part of the population has a different set of values that results in them

playing the second physical strategy. This speaks to an important part of multicultural ideas as to whether different cultures can coexist in the same population.
The proposition for the first model is as follows:

**Proposition 3a**:

If a population converges to an interior representational equilibrium then a different physical strategy is played in each representation iff the category combinations are (4v1) or (1v4).

Proof: See Appendix 1

The proposition for the second model is as follows:

**Proposition 3b:**

If a population converges to an interior representational equilibrium then a different physical strategy is played in each representation iff the category combinations are (4v1), (1v4), (2v4), (4v2), (2v1) or (1v2).

Proof: See Appendix 1

It follows that such a situation is possible- one *can* have uniform blocs where everyone with one set of values does one thing and everyone with another set of values does another. However, the likelihood of this happening depends crucially on whether a person is aware of the other person's choice of representation. Where they are aware then there are only two situations where this is likely to happen. Furthermore, perhaps ironically in a multicultural context, this situation frequently occurs where the values are diametrically opposite to each other.

Another question that is of importance is whether we can have different values within a population but with everyone playing the same physical strategy? This is of obvious importance in a multicultural society as the different values would not "matter" and potential clashes would be avoided. One would have different routes to the same conclusion. Unfortunately this possibility is not allowed within the current models:

For the first model:

**Proposition 4a:**

It is not possible to have a mixed representational equilibrium in a population where all members of the population play the same pure physical strategy.

Proof: See Appendix 1

For the second model:

**Proposition 4b**

It is not possible to have a mixed representational equilibrium in a population where all members of the population play the same pure physical strategy.

Proof: See Appendix 1

Even allowing for this, it is interesting to see whether we can have the opposite situation. In other words, can we have a situation where we have two sets of values coexisting and both strategies are being played by people holding either set of values? In such a situation we would have a "totally mixed" population in which representations and physical strategies can be played in any combination. Again, this would reflect a valued situation within a multicultural society.

For the first model:

**Proposition 5a**

A population will have a mixed equilibrium in representations and with a mixture of physical strategies played under both representations only if the game has a category combination (3v3).

Proof: See Appendix 1

For the second model:

**Proposition 5b**

It is not possible for a population to have a stable mixed equilibrium in representations and mixed equilibria in physical strategies in both subgames.

Proof: see Appendix 1

One can see here a clear divergence between the two models. In the first model a complete mixture of representations and physical strategies is allowed, so allowing a multicultural society to survive with both sides playing mixed physical strategies. However, this is not the case with the second model where mixed representations must be accompanied by at least one representation playing just one physical strategy.

Even allowing for this, in the first model, the playing of mixed strategies tends to be more where a population converges on one representation. This, again, is a result of the operation of the replicator dynamic on the truncated game. In most preference permutations the population will simply converge to a pure strategies equilibrium. In such a situation, the playing of multiple strategies within a population will be the result of playing a category 3 game in either or both of the symmetric subgames. Ironically, therefore, differences in values would not be a major force in driving behaviour and differing plays of physical strategies would most often emerge from everyone having the same set of values.

## 13. Discussion

The models outlined above are simple constructions aimed at creating basic frameworks in which we can discuss the origins and spread of values. As such they ignore many of the basic channels through which values are spread such as via parental nurturing, education, the mass media, law etc. The aim has been to strip the creation and spread of values down to its absolute basic framework. However, the models are different from many of the biology- influenced models put forward before. Values in this paper's models are not seen as unexplained "traits" passed down the generations or across a population but rather as a special type of preference which is characterised as being stable, context sensitive and widely held. The aim in this model is to take account of the psychological underpinnings of values in terms of emotion, interpretation and contextual sensitivity.

Given this, two simple evolutionary models were created that incorporated these issues into extensive form games. This allowed values to be split apart from physical actions while modelling the interactions between the two. Despite the simplicity of the models, some interesting results can be deduced from them. The propositions given above show what happens in a "classic" multicultural setting where there are more than one set of values coexisting in a population. It turns out that this setting is harder to achieve than may be thought and one such setting- where we have different sets of values but with everyone playing the same physical strategy, simply cannot exist as an equilibrium. Some other plausible settings such as those put forward in propositions 3a, 3b and 5a show that even populations with different strategies being played only coexist in particular ways when we have different values.

It is interesting to see how the two models vary from each other. The main difference that needed to be modelled was what happened when players could perceive each other's choice of representation (the first model) and when they could not (the second model). This required a considerably different modelling framework with the second model being considerably more "messy" than the first. However, many of the results are similar.

Propositions 1 and 2 relating to the second model are similar to what is discussed about the first model in the text. Proposition 3b is more permissive than proposition 3a, although the type of category combinations targeted are similar. Proposition 4a and 4b are identical so only 5a and 5b offer substantial differences in how much "mixing" is allowed.

This latter result difference does have interesting implications. It suggests that situations where one is familiar with another's values will have more mixing than situations where these values remain unknown. It suggests that a mixing of values within a culture will be far harder to sustain where there are a lot of anonymous interactions than where everyone knows each other. This would imply that multiculturalism in this sense would work best in societies where there is more knowledge of each other, more conservatism and less transfer of new ideas.

However, the most straightforward, and arguably the most interesting result, is that in many possible cases, the population will simply converge on one set of values. Indeed, even situations where there is a mixed equilibrium of physical strategies usually happen when the population only holds one set of values. It follows that, given the assumptions of the models, one would have a high expectation of the population converging on value uniformity with everyone sharing the same set of values. This undermines the notion that a multicultural society can survive without external "help".

This last result means that, if one is to set up a multicultural society with values surviving long-term beside each other then one needs to construct institutions or rules that *prevent* the operation of the learning mechanisms in the model. There are various ways in which this could be done. One radical method is through a process of "ghettoisation" where a population with one set of values is isolated from another majority population by physical location or by a set of rules. In this case learning would only take place within the isolated population and so there would be no transmission of values from the majority population. Even if parts of the population are not ghettoised then institutions such as firms may be able to influence a person's *work* values by partially isolating them from other influences.

Another possibility is that laws are passed and fines imposed that change the payoffs in the game. By changing the material payoffs one may hope to at least affect the utilities involved. In doing so, one could change the utilities within each representation to such an extent that they change to a different category that enables them to coexist. A similar process could occur if laws are passed that increase psychic costs to such a level that representations similarly change. Alternatively, laws may block off certain representations being transmitted- an example would be the laws against racial hatred.

An important way in which values may be sustained is through education since this provides a “short-cut” in learning values. This may be used in sustaining a multicultural society by training part of the population in one set of values. An example of this may be the creation of religious schools where religious values are incorporated into the curriculum. Another example may be where the school curriculum aims to replace all values within a population wholesale with a fixed set of values. An example of this may be the inculcation of scientific values such as objectivity and accuracy in school, replacing more primitive nonscientific ideas.

It follows that institutions and rules will form the values of the populations that are subject to them simply by intruding in the evolutionary process by which values come to be common within a population. Conversely, the dismantling of these rules and institutions could have a disruptive effect on the survival of certain values within a population. As we have seen, in the vast bulk of cases, there will be a tendency towards the assimilation of diverse values to one particular set.

The question of how norms can be transferred from one situation to another can also be solved by an extension of the current models. In novel situations, one has to decide how to interpret the situation and the actions of others in that situation. This means that one needs to find a suitable representation. The best way is to find a similar situation in one’s past history and to use the representation for that situation. In other words, one needs to use a common psychological process, that of analogy (c.f. Gentner et al 2001), to find a good fit. The mystery of why one would, for example, leave a tip in a restaurant that one has no intention of visiting again becomes plain. One leaves a tip because one has made an analogy with a representation of other restaurant situations and this restaurant is a good “fit” with that representation. The representation’s attached utilities then act as reasons for leaving a tip[11].

## 14. Conclusion

It is argued in this paper that the current models of values in economics are insufficient to explain many of the stylised facts that exist about values and, as a result, they cannot say much about the role of values in the economy and society. Part of the problem is

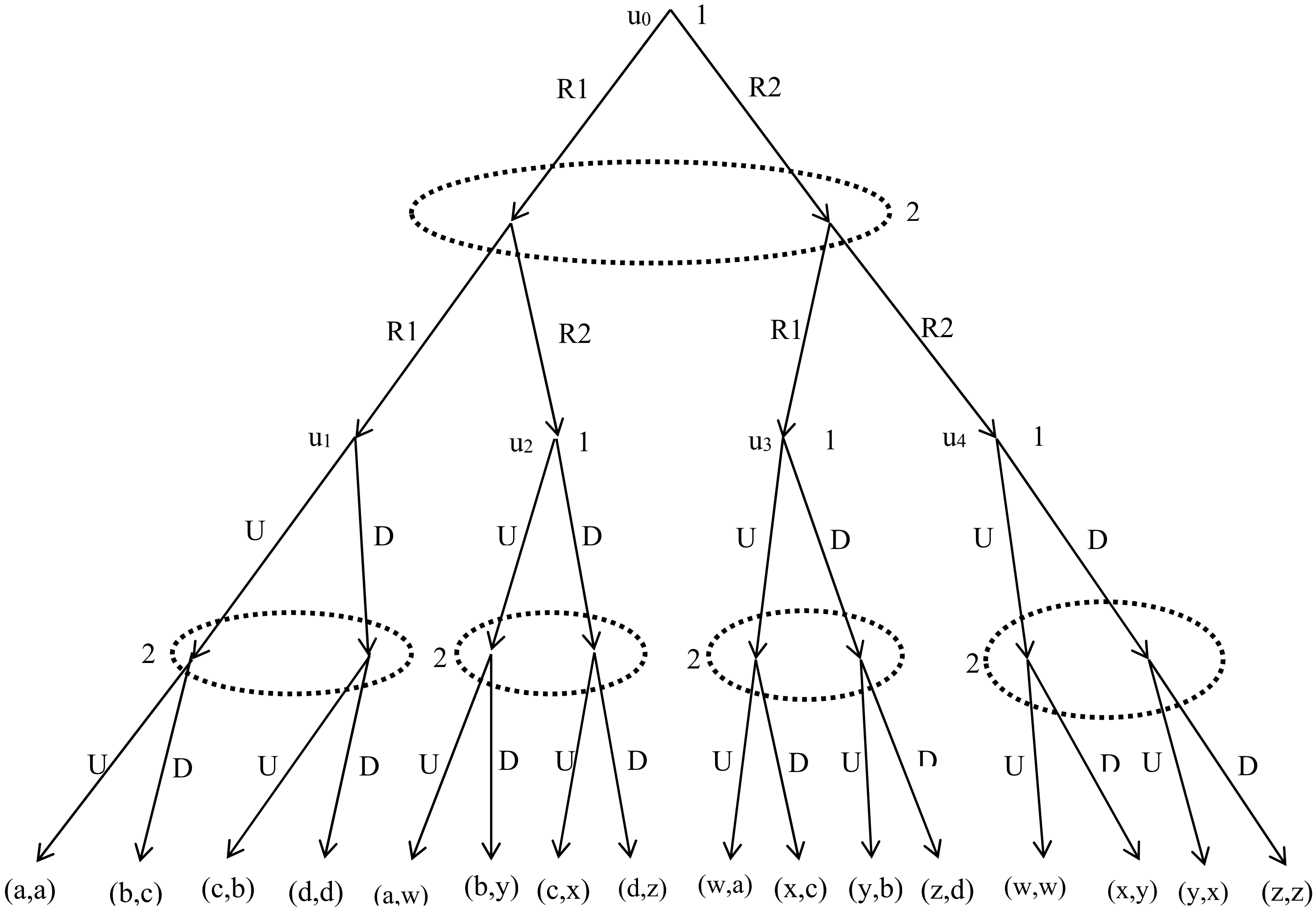
[11] Note that this is different from the explanation offered by Sugden (2005). For Sugden the reason for giving a tip is the result of legitimate expectations, which have normative content- one thinks that the giving of a tip is ethically right and so one does it in all situations. Here, normativity and the transmission of normative behaviour are separate. The transmission is the result of analogies made with representations of previous similar situations while normativity is the result of valuation.

that economists have ignored two ideas that have been widely used within the psychological and philosophical literature for a long period of time, namely the idea of a mental action and the idea of a representation.

When we take account of these factors then we can build two simple models that are surprisingly productive in terms of predictions and can shed light on current debates in multiculturalism adding structure and objectivity to a debate that is often highly politicised. Furthermore, we can use the model to solve more abstract difficulties relating to the creation of social norms and the question of the transmission of norms.

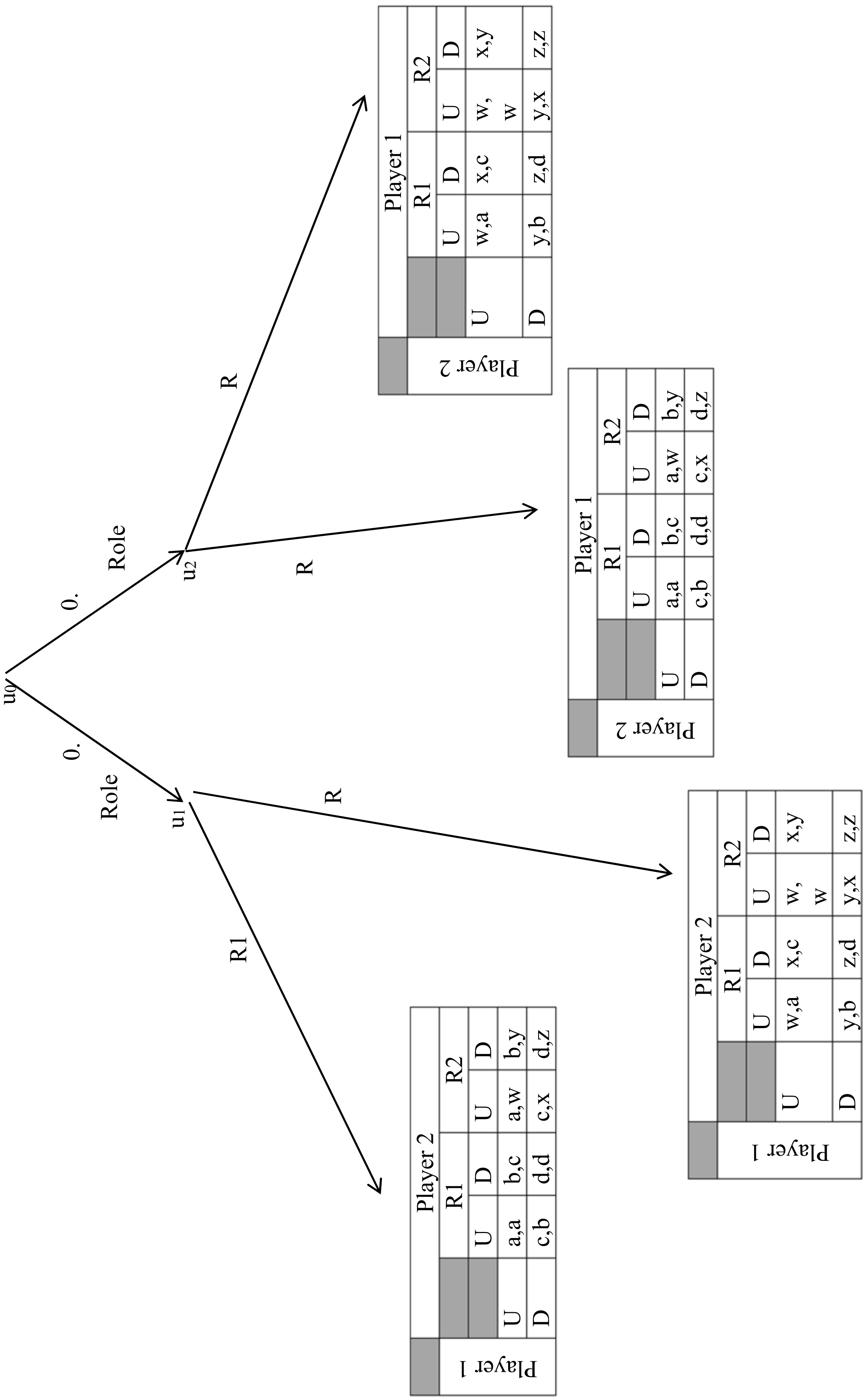
However, it should be noted that these models are still quite simple. A more realistic version would take account of the role of rules and institutions in value formation. Furthermore, the models are restricted to two strategy base games. There is no reason in principle why the models should not be expanded outwards to include more physical strategies and more representations. However, this may mean that some of the more specific conclusions in the paper may have to be modified.

Figure 1



u0 1
R1
R2
2
R1
R2
R1
R2
u1
u2 1
u3 1
u4 1
U D U D U D U D
2 2 2 2
U D U D U D U D U D U D U D U D
(a,a) (b,c) (c,b) (d,d) (a,w) (b,y) (c,x) (d,z) (w,a) (x,c) (y,b) (z,d) (w,w) (x,y) (y,x) (z,z)

Figure 2



u0
u1
u2
Role
0.
Role
0.
R1
R
R
R
Player 1
Player 2
R1
R2
U
D
a,a
b,c
a,w
b,y
c,b
d,d
c,x
d,z
w,a
x,c
w, w
x,y
y,b
z,d
y,x
z,z

Bibliography


Berner A. & Putterman L. (eds) (1998) “Economics Values and Organization” Cambridge University Press Cambridge

Binmore K (1988) “Modelling Rational Players II” Economics and Philosophy vol. 4 p. 9-55

Bisin A. & Verdier T.(2001) “The Economics of Cultural Transmission and the Dynamics of Preferences” Journal of Economic Theory 97, 298-319

Bowles S. & Gintis H. (2011) “A Cooperative Species: Human Reciprocity and its evolution” Princeton University Press Princeton

Cavalli-Sforza, L.L., Menozzi P., & Piazza A. (1981) Cultural Transmission and Evolution: A Quantitative Approach Princeton, Princeton University Press

Cressman R. (2003) “Evolutionary Dynamics and Extensive Form Games” MIT Press Cambridge Massachusetts

Davidson D. (1963) “Actions, Reasons and Causes” Symposium on Action The Journal of Philosophy 60 685-700

Dawkins R. (1976) The Selfish Gene Oxford Oxford University Press

Denzau A.T. & North D.C. (1994) “Shared mental models: Ideologies and institutions. Kyklos, Vol. 47, Issue 1

Fehr E. & Schmidt K.M. (1999) “A Theory of Fairness, Competition and Cooperation” Quarterly Journal of Economics 114 p. 817-868

Frank R. (1988) “Passions within Reason: The Strategic Role of Emotions” W.W. Norton & Company New York

Frey B.S. & Jegen R. (2001) “Motivation Crowding Theory” Journal of Economic Surveys 15 p. 589-611

Geach P.(1957) “Mental Acts: Their Content and their Objects” London Routledge & Kegan Paul.

Gentner D., Holyoak K.J. & Kokinov B.N. (2001) The Analogical Mind: Perspectives from Cognitive Science MIT Press Cambridge Massachusetts

Henrich J. & Boyd R. (2002) “On Modelling Cognition and Culture: Why Cultural Evolution does not Require Replication of Representations” Journal of Cognition and Culture 87-112

Henrich J., Boyd R., Bowles S., Camerer C., Fehr E., and Gintis H. (2004) “Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies” Oxford Oxford University Press

Inglehart R., Basañez M., and Moreno A. (1998) "Human values and beliefs :political, religious, sexual and economic norms in 43 societies. Findings from the 1990-1993 World Values Survey " Ann Arbor, Michigan. : University of Michigan Press

Jones M.K.(2016) "Representations and the Corruption of Goods" Dundee Discussion Papers in Economics no. 288

Jong, E. de, (2009) "Culture and economics :on values, economics and international business" London Routledge,.

Lewis D. (1969) "Convention: A philosophical Study" Wiley- Blackwell Oxford

Locke J. (1998) "An Essay Concerning Human Understanding" (first published 1689) Wordsworth editions Limited Hertfordshire

Mandler G. "Approaches to a Psychology of Value" in Hechter M., Nadel L. & Michod R.E (eds) (1993) "The Origin of Values" Walter de Gruyter Inc. New York

Mansbridge J. "Starting with nothing: on the impossibility of grounding norms in self-interest" in Berner A. & Putterman L. (eds) (1998) "Economics Values and Organization" Cambridge University Press Cambridge

Parekh B. (2006) "Rethinking Multiculturalism" (2nd Edition) Palgrave Macmillan Basingstoke Hampshire.

Proust J. (2001) "A Plea for Mental Acts" Synthese 129 105-128

Raz J. (2003) "The Practice of Value" The Berkeley Tanner Lectures Oxford University Press Oxford

Rokeach, M. (1973). The Nature of Human Values. New York: The Free Press.

Ross L. & Nisbett R.E. (1991) "The Person and the Situation: Perspectives of Social Psychology" McGraw- Hill Boston Massachussetts

Seligman C. & Katz A. "The Dynamics of Value Systems" in Seligman C., Olson J.M. & Zanna M. (eds) "The Psychology of Values" The Ontario symposium on Personality and Social Psychology vol. 8 (1996) Lawrence Erlbaum New Jersey.

Sen A. (1980) "Description as Choice" Oxford Economic Papers 32 p. 353-369

Simon H. (1959) "Theories of Decision-making in Economics and Behavioral Science" American Economic Review vol. 49 3 p. 253-283

Scharleman J., Eckel C. Kacelnik A. & Wilson R. (2001) "The Value of a Smile: Game Theory with a Human Face" Journal of Economic Psychology 22 p. 617-640.

Sperber D.(1996) "Explaining Culture: a naturalistic approach" Blackwell Oxford

Sugden, R. (2005) "The economics of rights, co-operation, and welfare" New York : Palgrave Macmillan.

Tabellini G. (2008) “The Scope of Cooperation: Values and Incentives” Quarterly Journal of Economics p. 905-950

Trevor-Roper H. “The Invention of tradition: The Highland tradition of Scotland” in Hobsbawm E. , Ranger T. (1983) (eds) The Invention of Tradition Cambridge University Press,

Neumann, J von and Morgenstern, O., (1953) Theory of Games and Economic Behavior. Princeton, NJ. Princeton University Press,.

Weibull J.W. (1995) “Evolutionary Game Theory” MIT Press Cambridge Massachussets.

## Appendix

Proofs for model 1:

**Proposition 3a**:

If a population converges to an interior representational equilibrium then a different physical strategy is played in each representation iff the category combinations are (4v1) or (1v4).

Proof:

First note that the (1v4) and (4v1) cases are identical so we only need to prove one to prove the other. We will prove the (4v1) case.

Assume subgame monotonicity and generic payoffs with (4v1) as the category combination.

For the symmetric subgame following $u_1$, use equation (2):

$$\dot{q}_U^{u_1} = P_{R1} q_U^{u_1}(1 - q_U^{u_1})\left[q_U^{u_1}(a-c) + \left(1 - q_U^{u_1}\right)(b-d)\right]$$

Assume that this is a category 4 game (i.e. a<c and b<d) so it must be the case that $q_U^{u_1} \to 0$ as t→∞.

For the symmetric subgame following $u_4$, use equation (5)

$$\dot{q}_U^{u_4} = (1 - P_{R1}) q_U^{u_4}(1 - q_U^{u_4})\left[q_U^{u_4}(w-y) + \left(1 - q_U^{u_4}\right)(x-z)\right]$$

Assume that this is a category 1 game (i.e. w>y and x>z) so it must be the case that $q_U^{u_4} \to 1$ as t→∞

For the linked asymmetric subgame, assuming the same payoffs and use equations (3) and (4)

$$\dot{q}_U^{u_2} = (1 - P_{R1}) q_U^{u_2}(1 - q_U^{u_2})\left[q_U^{u_3}(a-c) + \left(1 - q_U^{u_3}\right)(b-d)\right]$$

$$\dot{q}_U^{u_3} = P_{R1} q_U^{u_3}(1 - q_U^{u_3})\left[q_U^{u_2}(w-y) + \left(1 - q_U^{u_2}\right)(x-z)\right]$$

This results in $q_U^{u_2} \to 0$ as t→∞ and $q_U^{u_3} \to 1$ as t→∞. This results in the following truncated game payoff matrix:

| | $R_1$ | $R_2$ |
|---|---|---|
| $R_1$ | d,d | c,x |
| $R_2$ | x,c | w,w |

Assume that the population has a mixed equilibrium in representations. This means that d-x< 0 and w-c<0. By inspection of the payoff matrix, one can see that those members of the population that select Representation 1 will also only play D (since d and c only ever appear as utility outcomes of playing D) while those members of the population that select Representation 2 will also only play U (since x and w only ever appear as utility outcomes of playing U).

Now assume that we have only one strategy played in one representation and the other strategy only played in the second representation. We also assume that we have a mixed equilibrium in the truncated game.

None of the outcomes in the truncated game can be category 3 because the proposition excludes mixed action solutions by assumption. (2v3) is excluded anyway by assumption.

In each symmetric subgame there are a maximum of only two pure strategies equilibria that the population can converge on. This means there are two possible cases in a given symmetric subgame where the population converges to an equilibrium.

The symmetric subgame following $u_1$ is reached by playing $R_1$ while the symmetric subgame following $u_4$ is reached by playing $R_2$. Suppose that in the symmetric subgame following $u_1$, $q_U^{u_1} \rightarrow 0$ as t→∞. It follows that in the symmetric subgame following $u_4$, $q_U^{u_4} \rightarrow 1$ as t→∞ because everyone choosing $R_2$ must choose the opposite strategy by assumption. For there to be just one strategy played in one representation then for role 1 (playing $R_1$) in the linked asymmetric subgame $q_U^{u_2} \rightarrow 0$ as t→∞ as this is consistent in physical strategy with the subgame following $u_1$. For role 2 (playing $R_2$) $q_U^{u_3} \rightarrow 1$ as t→∞ as this is consistent in physical strategy with the subgame following $u_4$. Excluding category 3, this pattern of preferences can only occur with the category combination (4v1).

A similar argument establishes (1v4).■

**Proposition 4a:**

It is not possible to have a mixed representational equilibrium in a population where all members of the population play the same pure physical strategy.

Proof:

Proof by contradiction.

Assume that a mixed representational equilibrium exists. Also assume that we have generic payoffs. Fix an arbitrary strategy (since it is a symmetric game, we will choose U).

In the symmetric subgames: since category 3 is excluded and since we have assumed pure strategies so $q_U^{u_1} \to 1$ and $q_U^{u_4} \to 1$ as t→∞.

Assuming one has the same payoffs in the linked asymmetric subgame then $q_U^{u_2} \to 1$ and $q_U^{u_3} \to 1$ as t→∞.

This means that the truncated game matrix is as follows:

| | $R_1$ | $R_2$ |
|---|---|---|
| $R_1$ | a,a | a,w |
| $R_2$ | w,a | w,w |

Given generic payoffs, there can only be two possible preference orderings: a-w> 0 or a-w< 0. In the first case $P_{R1} \to 1$ as t→∞ with probability 1 while in the second case $P_{R1} \to 0$ as t→∞ with probability 1. This contradicts the assumption of a mixed representational equilibrium. A similar argument can be made for setting a physical strategy of D. Hence the proposition is true. ■

**Proposition 5a**

A population will have a mixed equilibrium in representations and with a mixture of physical strategies played under both representations only if the game has a category combination (3v3).

Proof:

Firstly note that it is not possible to have a situation where there are any mixed strategies in the asymmetric subgames because (2v3) and (3v2) are excluded by assumption.

Given this, there is only one way in which the proposition can hold i.e. if the symmetric subgames have mixed strategies equilibria. This means that $q_U^{u_1} \to \frac{d-b}{d-b+a-c}$ and $q_U^{u_4} \to \frac{z-x}{z-x+w-y}$ as t→∞. These ratios must both be positive and less than 1 so either the numerator and denominator are both positive or they are both negative. To be less than 1, the differences (z-x), (w-y), (d-b) and (a-c) must either be all positive or all negative. This suggests that they can only exist as equilibria under categories 2 or 3.

Examining equations 2 and 5:

$$\dot{q}_U^{u_1} = P_{R1} q_U^{u_1}(1 - q_U^{u_1})\left[q_U^{u_1}(a - c) + \left(1 - q_U^{u_1}\right)(b - d)\right]$$
$$\dot{q}_U^{u_4} = (1 - P_{R1}) q_U^{u_4}(1 - q_U^{u_4})\left[q_U^{u_4}(w - y) + \left(1 - q_U^{u_4}\right)(x - z)\right]$$

It can be seen that they only both converge on their interior equilibria if a<c, b>d,w<y and x>z. If this is the case then the category combination must be (3v3)■

Proofs for model 2:

Derivation of second model:

From the equations of motion for $t_i$ given in the text we can work out the equation for p. Since:

$$p = t_1 + t_2$$

$$q = \frac{t_1}{t_1 + t_2}$$

$$r = \frac{t_3}{t_3 + t_4}$$

$$\dot{p} = \dot{t}_1 + \dot{t}_2 = t_1X_1 + t_2X_2 - t_1t_1X_1 - t_1t_2X_2 - t_1t_3X_3 - t_1t_4X_4 - t_2t_1X_1 - t_2t_2X_2 - t_2t_3X_3 - t_2t_4X_4$$

$$\dot{p} = t_1X_1 + t_2X_2 - (t_1 + t_2)(t_1X_1 + t_2X_2 + t_3X_3 + t_4X_4)$$

Substituting:

$$\dot{p} = pqX_1 + p(1 - q)X_2 - p(pqX_1 + p(1 - q)X_2 + (1 - p)rX_3 + (1 - p(1 - r)X_4)$$

$$\dot{p} = p(1 - p)((qX_1 + (1 - q)X_2)) - (rX_3 + (1 - r)X_4))$$

The equations of motion for r and q can be derived in similar ways so only one derivation will be given here:

Differentiating:

$$\dot{q} = \frac{\dot{t}_1 t_2 - \dot{t}_2 t_1}{(t_1 + t_2)^2}$$

Substituting:

$$\dot{q} = \frac{t_2 t_1\big(X_1 - (t_1X_1 + t_2X_2 + t_3X_3 + t_4X_4)\big) - t_1 t_2(X_2 - (t_1X_1 + t_2X_2 + t_3X_3 + t_4X_4))}{(t_1 + t_2)^2}$$

Cancelling:

$$\dot{q} = \frac{t_1 t_2(X_1 - X_2)}{(t_1 + t_2)^2}$$

Substituting $X_1$ and $X_2$:

$$\dot{q} = \frac{t_1 t_2(((t_1 + t_3)a + (t_2 + t_4)b) - ((t_1 + t_3)c + (t_2 + t_4)d))}{(t_1 + t_2)^2}$$

Given that $p=t_1+t_2$ and the definitions of r and q above we can convert this to an equation in terms of p, q and r:

$$\dot{q} = \frac{qp\times(1-q)p\times\Big((qp+r(1-p))a+\big((1-q)p+(1-r)(1-p)\big)b-\Big((qp+r(1-p))c+\big((1-q)p+(1-r)(1-p)\big)d\Big)\Big)}{p^2}$$

Cancelling:

$$\dot{q} = q(1-q)\Big(\big(qp + r(1-p)\big)(a-c) + \big((1-q)p + (1-r)(1-p)\big)(b-d)\Big)$$

By a similar reasoning process:

$$\dot{r} = r(1-r)(\big(qp + r(1-p)\big)(w-y) + \big((1-q)p + (1-r)(1-p)\big)(x-z))$$

Propositions:

To prove the propositions for the second model in the paper we need the use of the following lemmas.

**Lemma 1**

Assume that populations converge on pure strategies equilibria in both subgames for a given role. For those that converge on the same pure strategy over both subgames there are no mixed representational equilibrium. For those that converge on different pure strategies over both subgames there is the possibility of converging either on a mixed representational equilibrium or a pure representational equilibrium.

Proof:

There are four possible combinations of pure strategy equilibria in the subgames: {q=1, r=1}, {q=0,r=0}, {q=1,r=0} and {q=0,r=1}. The former two correspond to the same pure strategy in each subgame and the latter two to different pure strategies in each subgame. Substituting the values of q and r into the equation of motion for p:

(1) {q=1, r=1}:

$$\dot{p} = p(1-p)(a-w)$$

(2) {q=0,r=0}

$$\dot{p} = p(1-p)(d-z)$$

(3) {q=0,r=1}

$$\dot{p} = p(1-p)(((1-p)c + pd) - ((1-p)w + px)))$$

(4) {q=1,r=0}

$$\dot{p} = p(1-p)((pa + (1-p)b) - (py + (1-p)z))$$

The first two equations depend on the values of (a-w) and (d-z) respectively. These are the payoffs from the games and so generically are constant either as a positive or a negative value. This implies that in these cases p will converge on 1 or 0 giving pure representations in either case.

The second pair of equations may result in p converging on 1 or 0. This would happen if, say, d>x or c<w for equation (3) or if b<z or a> y for equation (4). Alternatively p could converge on a mixed representation equilibrium at $p = \frac{c-w}{c-d-w+x}$ for equation (3) or $p = -\frac{b-z}{a-b-y+z}$ for equation (4). This would happen if x>d and c>w for equation (3) or y>a and b>z for equation 4.

**Lemma 2:**

All category combinations containing just the categories 1,2 or 4 will have a population that always converges on a pure strategy equilibrium.

Proof by cases:

Case 1: (1v1) and (4v4)
Examine the two equations of motion for the two subgames:

$$\dot{q} = q(1-q)\left(\left(qp + r(1-p)\right)(a-c) + \left((1-q)p + (1-r)(1-p)\right)(b-d)\right)$$

$$\dot{r} = r(1-r)(\left(qp + r(1-p)\right)(w-y) + \left((1-q)p + (1-r)(1-p)\right)(x-z))$$

If the category combination is (1v1) then (a-c)>0, (b-d)>0, (w-y)>0 and (x-z)>0. It follows that both r and q will converge to 1 i.e. pure strategy equilibria.

If the category combination is (4v4) then (a-c)<0, (b-d)<0, (w-y)<0 and (x-z)<0. It follows that both r and q will converge to 0.

Case 2: (2v2)

For (2v2) the category combinations are (a-c)>0, (b-d)<0, (w-y)>0 and (x-z)<0. Examining the equations of motion we can see that p increases positively with p and r when p is high and similarly with r when r is high. Conversely, p decreases with (1-p) and (1-r) when p is low and similarly with r when r is low.
The eigenvalues for the following equilibria are:
{p=1,q=1,r=1}: (w-a), (c-a), (y-w)

{p=0,q=1,r=1}: (a-w), (c-a), (y-w)
{p=1,q=0,r=0}: (z-d), (b-d), (x-z)
{p=0,q=0,r=0}: (d-z), (b-d), (x-z)
It can be seen that, ignoring the first value in each case, all of the eigenvalues are negative, while the first value depends on the relationship between the two subgames and is negative depending on the relationship between a and w or z and d.

Looking at the differentials of the equations of motion:

$$\frac{\partial \dot{q}}{\partial q} = (1-2q)\Big(\big(qp + r(1-p)\big)(a-c) + \big((1-q)p + (1-r)(1-p)\big)(b-d)\Big) + q(1-q)p\big((a-c)-(b-d)\big)$$

$$\frac{\partial \dot{r}}{\partial r} = (1-2r)\ \big(\big(qp + r(1-p)\big)(w-y) + ((1-q)p + (1-r)(1-p))(x-z)\big) + \ r(1-r)((1-p)((w-y)-(x-z))$$

It can be seen that interior rest points occur when :

$$\big(qp + r(1-p)\big)(a-c) + \big((1-q)p + (1-r)(1-p)\big)(b-d) = 0$$

and

$$\big(qp + r(1-p)\big)(w-y) + \big((1-q)p + (1-r)(1-p)\big)(x-z) = 0$$

Since (a-c)-(b-d) and (w-y)-(x-z) are both positive, it follows that some of the eigenvalues at the rest points must also be positive. It follows that there are no interior rest points that are stable.

Case 3: (1v2) (2v1) (4v2)(2v4):

For (1v2): (a-c)>0, (b-d)>0, (w-y)>0 and (x-z)<0
(2v1): (a-c)>0, (b-d)<0, (w-y)>0 and (x-z)>0
(4v2): (a-c)<0, (b-d)<0, (w-y)>0 and (x-z)<0
(2v4): (a-c)>0, (b-d)<0, (w-y)<0 and (x-z)<0

Examining the equations of motion:

$$\dot{q} = q(1-q)\Big(\big(qp + r(1-p)\big)(a-c) + \big((1-q)p + (1-r)(1-p)\big)(b-d)\Big)$$

$$\dot{r} = r(1-r)(\big(qp + r(1-p)\big)(w-y) + \big((1-q)p + (1-r)(1-p)\big)(x-z))$$

It can be seen that for category combination (1v2) q will converge to 1, for (2v1) r will converge to 1, for (4v2) q will converge to 0 and for (2v4) r will converge to 0.
This leaves the following equations of motion for each combination:

(1v2) $\dot{r} = r(1-r)(\big(p + r(1-p)\big)(w-y) + (1-r)(1-p)(x-z))$
(2v1) $\dot{q} = q(1-q)(\big(qp + (1-p)\big)(a-c) + \big((1-q)p\big)(b-d))$
(4v2) $\dot{r} = r(1-r)(\big(r(1-p)\big)(w-y) + \big(p + (1-r)(1-p)\big)(x-z))$
(2v4) $\dot{q} = q(1-q)(qp(a-c) + \big((1-q)p + (1-p)\big)(b-d))$

For each combination, the equation of motion describes motion with regard to category 2 payoffs. This means that there is motion towards the pure strategies equilibria in either direction and away from the mixed strategies rest points in the interior. For both equations for the motion of r, r moves positively in r and negatively in (1-r). Similarly for both equations for the motion of q, q moves positively in q and negatively in (1-q).

The following show all possible pure strategies equilibrium points and the corresponding eigenvalues with the category combinations that satisfy the last two eigenvalues:
{p=1,q=1,r=1}: (w-a), (c-a), (y-w); satisfied by (1v2) (2v1)
{p=0,q=1,r=1}: (a-w), (c-a), (y-w); satisfied by (1v2) (2v1)
{p=1,q=0,r=0}: (z-d), (b-d), (x-z); satisfied by (2v4) (4v2)
{p=0,q=0,r=0}: (d-z), (b-d), (x-z); satisfied by (2v4) (4v2)
{p=1,q=1,r=0}: (y-a), (c-a), (w-y); satisfied by (2v 4)
{p=0,q=0,r=1}: (c-w), (a-c), (y-w); satisfied by (4v2)
{p=1,q=0,r=1}: (x-d), (b-d), (z-x); satisfied by (2v1)
{p=0,q=1,r=0}: (b-z), (d-b), (x-z); satisfied by (1v2)

The first eigenvalue depends on the relationship of the payoffs between the two subgames.

The differentials of the equations of motion give the following results:

(1v2)

$$\frac{\partial \dot{r}}{\partial r} = (1-2r)\,\big(\big(p+r(1-p)\big)(w-y)+(1-r)(1-p)(x-z)\big)+\ r(1-r)((1-p)((w-y)-(x-z))$$

(2v1)

$$\frac{\partial \dot{q}}{\partial q} = (1-2q)\left((qp)(a-c)+\big((1-q)p+(1-p)\big)(b-d)\right)+q(1-q)p\big((a-c)-(b-d)\big)$$

(4v2)

$$\frac{\partial \dot{r}}{\partial r} = (1-2r)\,(\big(r(1-p)\big)(w-y)+((1-q)p+(1-r)(1-p))(x-z))+\ r(1-r)((1-p)((w-y)-(x-z))$$

(2v4)

$$\frac{\partial \dot{q}}{\partial q} = (1-2q)\left(\big(qp+(1-p)\big)(a-c)+\big((1-q)p\big)(b-d)\right)+q(1-q)p\big((a-c)-(b-d)\big)$$

The interior rest points are when the following holds:

(1v2): $\big(p+r(1-p)\big)(w-y)+(1-r)(1-p)(x-z)=0$

(2v1): $(qp)(a-c)+\big((1-q)p+(1-p)\big)(b-d)=0$

(4v2): $\big(r(1-p)\big)(w-y)+\big((1-q)p+(1-r)(1-p)\big)(x-z)=0$

(2v4): $\big(qp+(1-p)\big)(a-c)+\big((1-q)p\big)(b-d)=0$

Since (a-c)-(b-d) and (w-y)-(x-z) are both positive, it follows that the differentials will all be positive at that point meaning that the interior rest point is not stable.

Case 4: (1v4) (4v1)

(1v4): (a-c)>0, (b-d)>0, (w-y)<0 and (x-z)<0
(4v1): (a-c)<0, (b-d)<0, (w-y)>0 and (x-z)>0

Examine the equations of motion:

$$\dot{q}=q(1-q)\Big(\big(qp+r(1-p)\big)(a-c)+\big((1-q)p+(1-r)(1-p)\big)(b-d)\Big)$$

$$\dot{r}=r(1-r)(\big(qp+r(1-p)\big)(w-y)+\big((1-q)p+(1-r)(1-p)\big)(x-z))$$

It can be seen that for (4v1) q will always tend towards 0 while r will always tend towards 1. By contrast for (1v4) q will always tend towards 1 while r tends towards 0

**Lemma 3:**

Category combinations (3v1), (1v3), (3v2), (2v3), (3v4) and (4v3) all have a pure strategies equilibrium that is stable.
Proof by cases:
Case 1 (3v1) and (3v2)
Suppose {p=0,q=m,r=1} where the m represents the equilibrium probability of q.
The Jacobian has three eigenvalues:

$$T=ma+(1-m)c-w$$
$$R=-(m(a-c)+(1-m)(b-d))$$

and $S=y-w$

T is negative when c>w and depending on the values of m and a. R is negative in category 3 when $m<\frac{-(b-d)}{(a-c)-(b-d)}$ while S is negative in categories 1 and 2.
Case 2: (3v4) (and (3v2))
Suppose {p=0, q=m, r=0} where m represents the equilibrium probability of q.
The Jacobian has three eigenvalues:

$$T=mb+(1-m)d-z$$
$$R=(1-2m)(b-d)$$

and $S=x-z$

T is negative when b>z and depending on the values of m and d. R is negative when $m>0.5$ while S is negative in categories 2 and 4.
Case 3: (1v3) and (2v3)
Suppose {p=1, q=1, r=k} where k represents the equilibrium probability of r.

The Jacobian has three eigenvalues:

$$T = kw + (1-k)y - a$$

$$R = c - a$$

and

$$S = (1-2k)(w-y)$$

T is negative when y>a and depending on the values of k and w. S is negative when $k < 0.5$ while R is negative in categories 1 and 2.

Case 4: (2v3) and (4v3)
Suppose {p=1, q=0, r=k} where k represents the equilibrium probability of r.
The Jacobian has three eigenvalues:

$$T = kx + (1-k)z - d$$

$$R = b - d$$

and

$$S = (1-2k)(x-z)$$

T is negative when z>d and depending on the values of k and x. S is negative when $k > 0.5$ while R is negative in categories 2 and 4.

**Lemma 4**

The surfaces of mixed strategy rest points in the two subgames do not have a curve of intersection within the unit probability space.

The two surfaces are described implicitly by the two relationships below

$$\Big(\big(qp + r(1-p)\big)(a-c) + \big((1-q)p + (1-r)(1-p)\big)(b-d)\Big) = 0$$

$$\Big(\big(qp + r(1-p)\big)(w-y) + \big((1-q)p + (1-r)(1-p)\big)(x-z)\Big) = 0$$

Making p the dependent variable of both:

$$p = \frac{-(r(a-c) + (1-r)(b-d))}{(q-r)((a-c)-(b-d))}$$

$$p = \frac{-(r(w-y) + (1-r)(x-z))}{(q-r)((w-y)-(x-z))}$$

Setting ps equal to each other, cross- multiplying and cancelling, we end up with:

$$(b-d)(w-y) - (a-c)(x-z) = 0$$

This implies that the two surfaces can only intersect if the above relationship between the constants is fulfilled. This implies that in general there is no intersection between the two surfaces.

**Proposition 1:**

a) Most equilibria in category combinations containing just categories 1, 2 and 4 are situations where just one representation is being played by the entire population.
b) All category combinations containing category 3 and one of categories 1,2 and 4 have an equilibrium where just one representation is played by the entire population.

Proof:
a) This follows from Lemma 1
b) This follows from the cases given in lemma 3

**Proposition 2:**

a) Most equilibria in category combinations containing just categories 1, 2 and 4 are situations where just one physical strategy is being played by the entire population.
b) All category combinations containing category 3 and one of categories 1,2 and 4 have an equilibrium where just one physical strategy is played by the entire population.

Proof:
a) From lemma 2 it can be seen that all category combinations containing categories 1,2 and 4 converge on pure strategies equilibrium in the subgames. From lemma 1 it can be seen that for pure strategies equilibria in subgames, the representations converge on mixed representational equilibrium only when {q=1,r=0} or {q=0, r=1} and even in these cases, most payoff orderings will end up with convergence on pure representational equilibrium. Hence the majority of representational equilibria are pure representational equilibria. Since each representation means that all play is within one subgame then this means that only one strategy will be played by the entire population.
b) This follows directly from lemma 3 where all cases involved a pure representational equilibrium that allowed play only in the subgame with pure strategies equilibrium.

**Proposition 3b:**

If a population converges to an interior representational equilibrium then a different pure strategy is played in each representation iff the category combinations are (4v1), (1v4), (2v4), (4v2), (2v1) or (1v2).

If:
Proof by cases:
Case 1: (4v1) and (1v4)
From lemma 2 case 4 it can be seen that (4v1) converges so that q=0, r=1 while for (1v4) the population converges to q=1, r=0. From lemma 1 it can be seen that these values allow for mixed representational equilibrium.
Case 2: (2v4), (4v2), (2v1), (1v2)
Lemma 2 case 3 show that in all these cases convergence on either q=0, r=1 or q=1, r=0 is possible. Lemma 1 demonstrates that these values allow for a mixed representational equilibrium.

Only if:
Assume an interior representational equilibrium. By lemma 1 this can only take place either when q=0, r=1 or q=1, r=0 which is definitionally when different pure strategies are played in different representations.
The eigenvalues for either situation are as follows:
{p=1,q=1,r=0}: (y-a), (c-a), (w-y);
{p=0,q=0,r=1}: (c-w), (a-c), (y-w);
{p=1,q=0,r=1}: (x-d), (b-d), (z-x);

{p=0,q=1,r=0}: (b-z), (d-b), (x-z);
The only category combinations that satisfy the two right hand side eigenvalues in any of these four situations are (4v1), (1v4), (2v4), (4v2), (2v1) or (1v2).

**Proposition 4b:**

It is not possible to have a mixed representational equilibrium in a population where all members of the population play the same pure physical strategy.
Proof:
There are two cases where the same pure strategy is played in both subgames: either q=1 and r=1 or q=0 and r=0. The equations of motion for p in each of these cases is shown below:

(1) {q=1, r=1}:

$$\dot{p} = p(1-p)(a-w)$$

(2) {q=0,r=0}

$$\dot{p} = p(1-p)(d-z)$$

Inspection of these equations shows that, in general, p will only converge on p=1 or p=0. Hence a mixed representational equilibrium is not possible.

**Proposition 5b**

It is not possible for a population to have a stable mixed equilibrium in representations and mixed equilibria in physical strategies in both subgames.
Proof:
This follows directly from lemma 5. If the two surfaces of mixed strategies rest points do not intersect then there is no point where an interior equilibrium can be stable.